3/9/93

Course 4.902: Pattern Recognition and Analysis
Solutions to Problem Set 2

Solution to Exercise 1: (5 points)

An $m \times n$ matrix $\mathbf{A}$ with real elements a_{ij} is used to form another matrix $\mathbf{B} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. Let b_{ij} denote the element in the i^{th} row and the j^{th} column of $\mathbf{B}$. Using the rule for matrix multiplication and the definition of transpose, we have

$$\begin{aligned} b_{ij} &= \sum_{k=1}^{n} a_{ik} a_{jk} \\ &= \sum_{k=1}^{n} a_{jk} a_{ik} \\ &= b_{ji} \end{aligned}$$

so that $\mathbf{B}$ is symmetric.

To show that $\mathbf{B}$ is positive semidefinite, we let $\mathbf{x}$ be an arbitrary nonzero m-dimensional vector, and consider the the quadratic form $\mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}$:

$$\begin{aligned} \mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x} &= \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{x} \\ &= (\mathbf{A}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{x} \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y}, \end{aligned}$$

where $\mathbf{y} = \mathbf{A}^{\mathrm{T}}\mathbf{x}$. Note that $\mathbf{y}^{\mathrm{T}}\mathbf{y} = |\mathbf{y}|^{\mathbf{2}} \geq \mathbf{0}$, so that $\mathbf{B}$ is positive semidefinite as desired.

Solution to Exercise 2: (5 points)

Since the two classes of data have some overlap, any threshold will result in classification errors. The total number of errors for the three given thresholds are:

Threshold	Number of Errors
6.7	1
6.9	2
7.1	1

Although we can get a rough idea of what the decision boundary should be based on minimizing the number of errors, we cannot fine-tune it in this way, since 6.7 and 7.1 are equally good, while their midpoint is worse than either. To fine-tune the boundary, we would either have to obtain a much larger number of sample points, or else we would have to do some sort of smoothing of the sample distributions.

Solution to Exercise 3: (5 points)

The answer is (b), nonsingular.

Solution to Exercise 4: (5 points)

The answer is (d), a possible covariance matrix. This is because it is symmetric (which was given) and has strictly positive eigenvalues.

Solution to Exercise 5: (5 points)

The answer is (c), square. The diagonal elements being positive does not tell us anything about the signs of the eigenvalues, so the most we can say is that the matrix is square (since it is symmetric).

Solution to Exercise 6: (5 points)

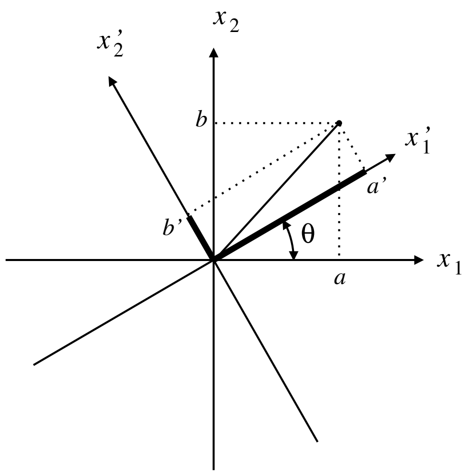
Let $\mathbf{x}$ be a random N-vector with mean $\mathbf{m_x}$, correlation matrix $\mathbf{R}$, and covariance matrix $\mathbf{K}$. To show that $\mathbf{R}$ is positive semidefinite, let $\mathbf{a}$ be an arbitrary nonzero N-vector, and consider the quadratic form

$$\begin{aligned}
\mathbf{a}^{\mathrm{T}}\mathbf{R}\mathbf{a} &= \mathbf{a}^{\mathrm{T}} E[\mathbf{x}\mathbf{x}^{\mathrm{T}}]\mathbf{a} \\
&= E[\mathbf{a}^{\mathrm{T}}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{a}] \\
&= E[(\mathbf{x}^{\mathrm{T}}\mathbf{a})^{\mathrm{T}}\mathbf{x}^{\mathrm{T}}\mathbf{a}] \\
&= E[(\mathbf{x}^{\mathrm{T}}\mathbf{a})^2] \\
&\geq 0,
\end{aligned}$$

so that $\mathbf{R}$ is positive semidefinite. To see that $\mathbf{K}$ is also positive semidefinite, note that $\mathbf{K}$ is the correlation matrix of the random vector $\mathbf{y} = \mathbf{x} - \mathbf{m_x}$.

Solution to Problem 1: (15 points)

Let $\mathbf{p}' = [a' \;\; b']^{\mathrm{T}}$ be the coordinate vector in the primed system. The coordinate a' is the projection of $\mathbf{p}$ onto the x_1' axis, and b' is the projection onto the x_2' axis, as indicated by the bold lines in the diagram below.



Working in the unprimed system, the unit vector in the x_1' direction is $[\cos\theta \;\; \sin\theta]^{\mathrm{T}}$, so that

$$a' = [\cos\theta \;\; \sin\theta]\begin{bmatrix} a \\ b \end{bmatrix}$$

Similarly, the unit vector in the x_2' direction is $[-\sin\theta \;\; \cos\theta]^{\mathrm{T}}$, so that

$$b' = [-\sin\theta \;\; \cos\theta]\begin{bmatrix} a \\ b \end{bmatrix}$$

In matrix form,

$$\mathbf{p}' = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\mathbf{p}$$

So that

$$\mathbf{A} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

is the desired rotation matrix.

To answer part (b), we can use results given in the book. In particular, Eq. (4.49) gives

$$\mathbf{K_y} = \mathbf{A}\mathbf{K_x}\mathbf{A}^T$$

and, since $\mathbf{A}$ is orthonormal, Eq. (4.50c) gives

$$\det(\mathbf{K_y}) = \det(\mathbf{K_x}).$$

Now let

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{K_x}.$$

A straightforward computation gives the two diagonal elements of $\mathbf{K_y}$ as

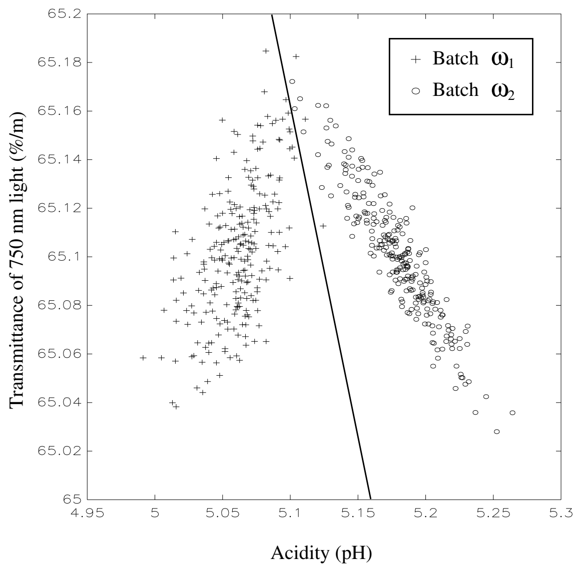
$$\begin{aligned} \text{top left element} &= a\cos^2\theta + b\cos\theta\sin\theta + c\cos\theta\sin\theta + d\sin^2\theta \\ \text{bottom right element} &= a\sin^2\theta - b\cos\theta\sin\theta - c\cos\theta\sin\theta + d\cos^2\theta, \end{aligned}$$

so that the trace of $\mathbf{K_y}$ is

$$\begin{aligned} \mathrm{tr}(\mathbf{K_y}) &= a\,[\cos^2\theta + \sin^2\theta] + d\,[\sin^2\theta + \cos^2\theta] \\ &= a + d \\ &= \mathrm{tr}(\mathbf{K_x}). \end{aligned}$$

Solution to Problem 2: (15 points)

a) A suitable line is indicated in the diagram below. The equation of the line is approximately $y = -2.5x + 78$. Any other line that divides the classes reasonably well would also do.



b) There is no one right answer here. The number of samples in each class is 250, and the number of classification errors using the line in part (a) is 3. Since 3/500 is much larger than 10^{-4}, it is unlikely that the error probability goal will be met.

To reduce the error probability, some sort of averaging could be used. For instance, rather than base the classification on testing a single pill, all 250 pills from the selected bottle could be tested, and the results averaged. The conditional variance in each dimension will be reduced by a factor of 250, so that the error probability will be dramatically reduced. To get a rough idea of whether or not the resulting error probability will meet the goal, we can model the class conditional densities as Gaussian. The class conditional density of the projection of a data point onto a line perpendicular to the decision boundary will also be Gaussian, because of one of the properties of Gaussian random vectors. Judging by the error probability before averaging, we can say that the decision boundary is definitely more than one standard deviation from each of the two class conditional means (in fact, it appears to be much more than this). After averaging, the distance of the decision boundary from the means is effectively multiplied by $\sqrt{250} = 16$. Therefore, the error probability for each class is less than the area under the tail of the unit Gaussian curve beginning at 16 standard deviations from the mean, which is an extremely small number, much smaller than 10^{-4}. Thus, the error probability goal will easily be met.

A sequential test may not be appropriate here, since we are not trying to minimize the number of pills used in this case – we might as well use the whole bottle, which easily lets us achieve the error-probability goal.

c) It is clear from the scatter plot that the acidity measurement is providing most of the information. The means appear to be well separated in acidity, but not in IR transmittance, while the variances are comparable. Hence, the acidity measurements should be chosen as the single type of measurement. This amounts to requiring that the line in the above diagram be vertical. Using the same reasoning as in part (b), it can be seen that again averaging over the entire contents of the bottle (250 pills) will allow the error-probability goal to be met easily.

Solution to Problem 3: (20 points)

The conditional densities are

$$p_{y|\omega_1}(\hat{y}|\omega_1) = k_1 \exp(-\hat{y}^2/6)$$

given class ω_1, and

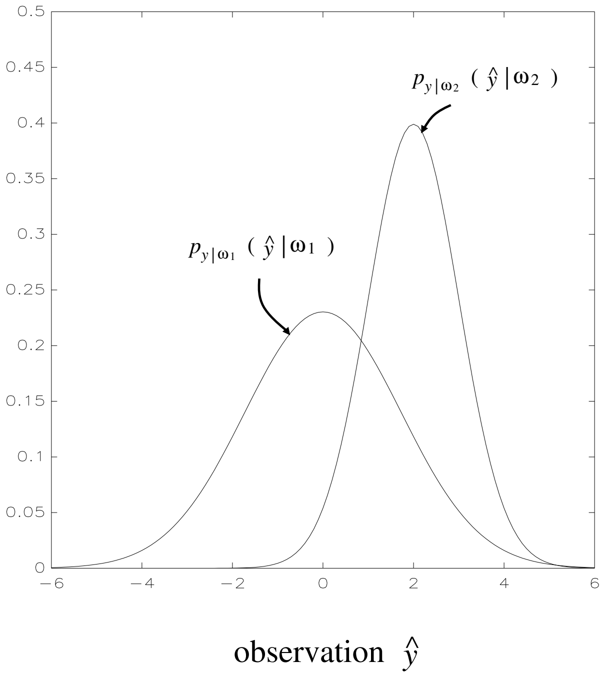
$$p_{y|\omega_2}(\hat{y}|\omega_2) = k_2 \exp(-(\hat{y}-2)^2/2)$$

given class ω_2.

a) The values for k_1 and k_2 are

$$k_1 = 1/\sqrt{6\pi}; \qquad k_1 = 1/\sqrt{2\pi}$$

which can be obtained by noting (from the exponents) that the variances are 3 and 1 for classes ω_1 and ω_2 respectively. The graphs of these conditional densities are given below.



b) Assuming equal priors and a cost assignment of

$$C_{11} = C_{22} = 0; \qquad C_{12} = 1, \quad C_{21} = \sqrt{3},$$

the expression for Bayes risk (Eq. (3.15) in the text) becomes

$$\mathcal{R} = \frac{\sqrt{3}}{2}\epsilon_1 + \epsilon_2/2$$

where

$$\epsilon_1 = 1/\sqrt{6\pi}\int_{R_2} \exp(-y^2/6)dy$$

and

$$\epsilon_2 = 1/\sqrt{2\pi}\int_{R_1} \exp(-(y-2)^2/2)dy$$

and where R_i denotes the region in which it will be decided that ω_i occurred.

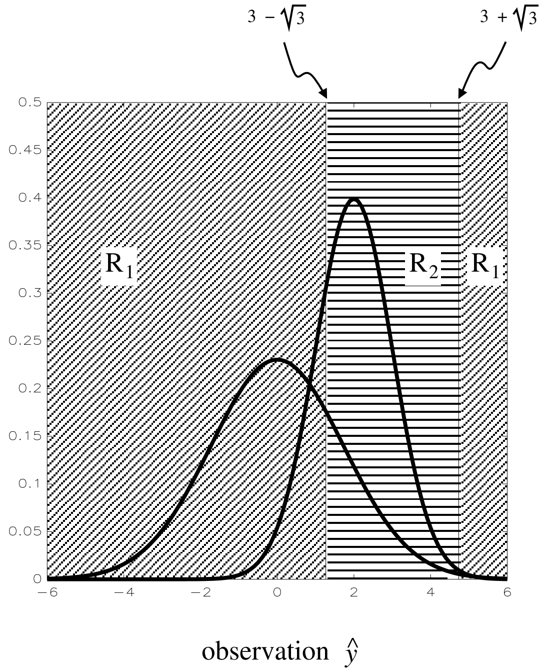
c) To minimize Bayes risk, the regions are defined by the following likelihood ratio test (see Eq. (3.14b) in the text):

$$\frac{1/\sqrt{6\pi}\exp(-\hat{y}^2/6)}{1/\sqrt{2\pi}\exp(-(\hat{y}-2)^2/2)} \quad \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \quad \frac{1}{\sqrt{3}}$$

after cancelling the $1/\sqrt{3}$ from both sides, taking the logarithm, and simplifying, the test becomes

$$\hat{y}^2 - 6\hat{y} + 6 \quad \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \quad 0.$$

The two thresholds are therefore $3-\sqrt{3}$ and $3+\sqrt{3}$, and the regions are as indicated below.



d) If x is a Gaussian random variable with mean μ and standard deviation σ, then the random variable $z = (x-\mu)/\sigma$ is a standard (zero-mean, unit-variance) Gaussian random variable. Let $\Phi(z)$ denote the cumulative distribution function of a standard Gaussian random variable. (Tables of Φ are widely available.) Then

$$\begin{aligned}\epsilon_1 &= 1/\sqrt{6\pi}\int_{R_2}\exp(-y^2/6)dy \\ &= \Phi(\frac{3+\sqrt{3}}{\sqrt{3}}) - \Phi(\frac{3-\sqrt{3}}{\sqrt{3}}) \\ &\approx 0.229,\end{aligned}$$

and

$$\begin{aligned}\epsilon_2 &= 1/\sqrt{2\pi}\int_{R_1} \exp(-(y-2)^2/2)dy \\ &= \Phi(3-\sqrt{3}-2) + [1-\Phi(3+\sqrt{3}-2)] \\ &\approx 0.235.\end{aligned}$$

Using the expression for $\mathcal{R}$ from part (b),

$$\begin{aligned}\mathcal{R} &\approx \frac{\sqrt{3}}{2}(0.229) + 0.235/2 \\ &= 0.316.\end{aligned}$$

Solution to Problem 4: (20 points)

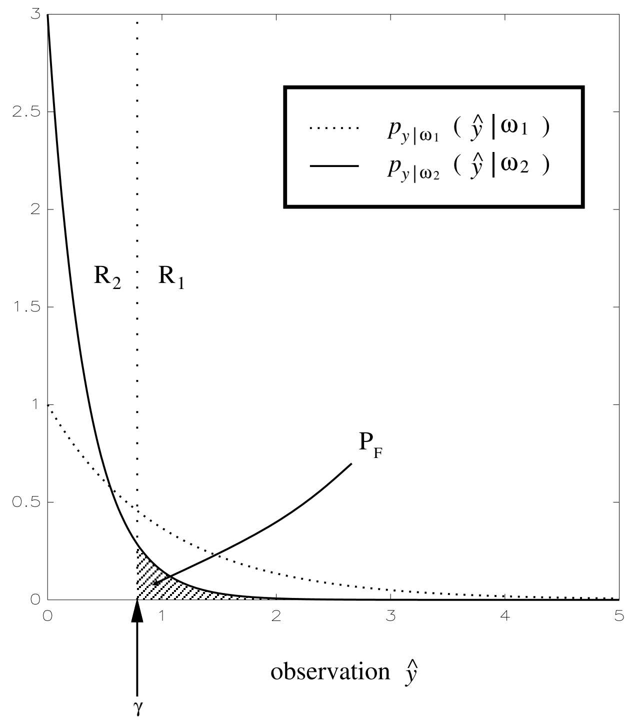
Since we are trying to maximize the detection probability P_D while constraining the false-alarm probability P_F to be no greater than a given amount, the problem is a Neyman-Pearson hypothesis testing problem. We know that the optimum test will be a likelihood ratio test (LRT), but we don't yet know what the threshold should be (we must determine it). For $\hat{y} > 0$ (which is the case of interest), the LRT is

$$\frac{1}{3}\frac{e^{-\hat{y}}}{e^{-3\hat{y}}} \quad \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \quad \lambda$$

for some unknown λ, which can be rewritten by taking the logarithm and simplifying as

$$\hat{y} \quad \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \quad \gamma,$$

where γ is to be determined by constraining P_F. To understand the role of γ, it is instructive to consider the following diagram, where P_F is indicated as the shaded area.



Solving for γ,

$$
\begin{aligned}
P_F &= \int_{\gamma}^{\infty} 3e^{-3y} dy \\
&= e^{-3\gamma} \\
&= 0.05
\end{aligned}
$$

so that $\gamma = -\ln(0.05)/3 \approx 1.0$. Finally, to solve for P_D, we use

$$
\begin{aligned}
P_D &= \int_{\gamma}^{\infty} e^{-y} dy \\
&\approx e^{-1} \\
&\approx 0.37.
\end{aligned}
$$

Solution to Puzzle: (0 points)

(Petersburg Paradox or St. Petersburg Paradox)

If the bank's resources are unlimited, then the expected value of the winnings is infinite, which is the "paradox." (Actually strictly speaking, it is undefined, but the formula gives infinity.)

The remainder of the problem involves a modification to the game that may make it seem less paradoxical. If the bank's resources are restricted so that it will never pay out more than 2^M, then the expected value of the winnings W becomes

$$
\begin{aligned}
E[W] &= \sum_{i=1}^{M} 2^{-i} 2^i + 2^M \sum_{i=M+1}^{\infty} 2^{-i} \\
&= M + 2^M 2^{-M} \sum_{i=1}^{\infty} 2^{-i} \\
&= M + 1.
\end{aligned}
$$

The probability that the player will make a profit, assuming an entry fee of $M+1$, is the probability that n will exceed n_0, where

$$n_0 = \lceil \log_2(M+1) \rceil,$$

where $\lceil x \rceil$ denotes the smallest integer not less than x. Note that as the bank's resources increase, the chances of the player making a profit decrease.